LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-TR-658996

When did the Haiti cholera outbreak begin?

S. P. Velsko

August 18, 2014

Disclaimer

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

When did the Haiti cholera outbreak begin?

Stephan P. Velsko
Lawrence Livermore National Laboratory
July 15, 2014

Summary

The available epidemiological and molecular clock evidence suggests that there is a considerable chance that cholera was introduced into Haiti prior to the arrival of Nepalese troops on October 8th, 2010. Thus, scientific support for widely expressed opinions about the timing of the outbreak may be weaker than first impressions might suggest.

Introduction

When did the Haiti cholera outbreak of 2010 begin? The widely accepted narrative is that cholera was brought to Haiti by a contingent of Nepalese soldiers who were part of the United Nations peacekeeping forces [1,2]. Cholera was first confirmed in cases that appeared shortly after the troops' arrival in the geographic vicinity of the Nepalese contingent's camp near Mirebalais in central Haiti. This fact is often cited as one of the critical pieces of evidence for the culpability of the Nepalese [3-5].

However, the certainty of this narrative must be tempered by the knowledge that a disease can be present in a region long before physicians recognize it. For example sporadic cases of HIV/AIDS almost certainly occurred in the United States years before the outbreak was recognized in 1981[6]. In the case of cholera, the predominance of mild or asymptomatic cases [7], and the high incidence of other diarrheal diseases in Haiti [8] provide plausible conditions for belated recognition. To be precise, it seems possible that cholera introduction and limited spreading could occur for many days before a case with easily recognized "classical" cholera symptoms appears.

From a forensic perspective, the critical question is how likely could it be that cholera was introduced into Haiti *before* the arrival of the Nepalese contingent? To answer this question, there are two main areas of scientific evidence to consider. First, findings regarding the incubation time and case severity distribution of cholera can be used to deduce the probability that a victim with a known date of onset was infected a certain number of days prior. Second, "molecular clock" analysis can estimate dates for the initiation of clonal expansion of the outbreak strain. This document summarizes what published information in these two areas allows us to infer about when the Haiti outbreak began.

Background information

One of the potential pitfalls in any forensic investigation is to consider only what the available evidence says about the "favored hypothesis" and not what it says about alternative hypotheses. The analysis in this report implicitly considers the

alternative hypothesis that someone other than the Nepalese peacekeeping contingent introduced cholera into Haiti. The case against the Nepalese contingent is notable for the absence of direct evidence of cholera in the Nepalese camp. No evidence has been offered that any Nepalese soldiers were ill with cholera, and no molecular evidence of *Vibrio cholerae* in the camp sanitary system or in the discharges that evidently contaminated the nearby Artibonite river system has been forthcoming. Alternative routes for cholera introduction are other travelers visiting the Mirebalais area prior to the outbreak. Possible examples include:

- Nepalese government officials visiting in preparation for the troop transfer
- United Nations officials who visited Nepal then Haiti in preparation for the troop transfer.
- Nepalese citizens with family connections to troops at the camp

In the analysis that follows we will show that the epidemiological evidence provides only mild support for the infection of a widely cited "index case" just after the arrival of the Nepalese contingent, while the molecular clock analysis could be argued to support an infection date *prior* to arrival.

We will take the following as facts pertaining to the epidemiological investigations carried out in Haiti after the outbreak:

(1) The Nepalese contingent began arriving at the Mirebalais camp on October 8th, and troop transfer was evidently complete by the 10th [9].

(2) The first suspected case of cholera was reported to the Haitian Public Health organization (MSPP) on October 18th [10], and the first positive test for cholera was obtained on October 19th [11].

(3) At least one presumed victim, cited in several places as the "index case" showed initial symptoms of cholera on October 12th [12]. However this patient did not go to a hospital, and the diagnosis was never laboratory-confirmed. A more recent review article [13] seems to imply that this was a confirmed cholera case, but no authority is cited.

(4) There was a steady "background" stream of between 5 and 10 cases per day of severe diarrhea into the Mirebalais Government Hospital from the 8th to the 18th of October[9]. Published data does not cover dates prior to that, but presumably any testing that occurred was negative for cholera [14]. (It would be informative to know if any regular testing for cholera was done for diarrhea cases presenting at Mirebalais hospital, and whether negative testing results are on record for the period prior to the first positive test result.) Public health surveillance for cholera

may have been influenced by the CDC assessment published in February, 2010 that while diarrheal disease outbreaks were expected, cholera was not likely [15].

Epidemiological Analysis

The general question to be addressed is: what is the probability that cholera was actually present prior to the arrival of the Nepalese contingent? From an epidemiological perspective, we can ask the more specific question: if the earliest onset of cholera symptoms was observed in the "index case" on the 12th, what is the probability that the victim was infected on the 11th, 10th, 9th, 8th, etc.? We have developed a simple stochastic modeling approach to answering this question.

First, from epidemiological data on the delay between infection with cholera and the appearance of symptoms (called the incubation time distribution) we can construct the probability that, if N patients were infected on day 0, then the first patient would seek medical attention on day 0+K (thereby allowing physicians to detect the presence of cholera). The details of the model are contained in appendix A. Next, we consider the probability that, if a person appears at a hospital with severe cholera symptoms on day 0, he was one of N persons infected on day –K. In appendix B we show that this posterior probability is simply the "mirror" of the distribution derived in Appendix A.

Calculations of these probabilities depend on the fraction of cholera cases that are mild or asymptomatic. A variety of estimates have been provided for this fraction, which is often taken to be around 75% [16-18]. However, a 25% probability of hospital-seeking (P_S) may be an overestimate for Haiti, as suggested by the "index case" himself.

A recent review by Azman et. al. provides up-to-date data on the incubation time distribution [19]. This review makes a highly relevant observation that the incubation time distribution for the O1 El Tor strain that affected Haiti has a very long right-hand tail. Thus, a long delay between infection and the development of symptoms is experienced by a relatively large fraction of patients.

Table 1. Estimated posterior probability of infection date, assuming the likelihood of hospital-seeking P_S = 25% and various initial infection numbers.

N	P(≥ Oct. 8th)	P(< Oct. 8th)	Likelihood Ratio
1	0.78	0.22	3.5
3	0.83	0.17	4.9
10	0.94	0.06	16
30	1.00	0.00	∞

The results in Table 1 indicate that, if there were only a few initial infections, the probability that infection of the "index case" could have occurred prior to October 8th is around 20%. Concomitantly, likelihood ratios supporting the hypothesis of infection ≥ Oct. 8th are less than 10, generally considered a weak level of support

[20]. If the likelihood of hospital-seeking is lower than 25%, then P(< Oct. 8th) increases. For example, if P_S = 10%, then P(< Oct. 8th) ≈ 0.14 for N = 10. Thus, published epidemiological information on the Haiti outbreak leaves considerable room for an earlier start date.

Molecular Clock Analysis

Katz et. al. have published a molecular clock analysis of the start of the Haiti outbreak [21]. This analysis relies on finding the common ancestor of a large number of sequences of Vibrio cholera isolated from victims of the Haiti outbreak, and assuming that the accumulation of mutations in a sequence depends linearly on the rate of expansion of the clonal population of sequences. Given the known dates at which the isolates were obtained, and the overall mutation rate of the strain, one can extrapolate back to the initial date of infection, assumed to coincide with the age of the common ancestor sequence.

The molecular clock analysis estimated a most recent common ancestor date of September 28, 2010. The "95% credibility interval" for the estimate was July 23rd to October 17th. The authors note:

> "The credibility interval encompasses the date that the Nepalese soldiers arrived in Haiti …, as well as the first reported hospitalization of a cholera case … (although an earlier fatal case with an onset date of 12 October may have been the index case)." … "Our results suggest that a population genomic approach *can be very powerful in delimiting the time frame of an outbreak*" ([21], page 6) [Emphasis added]

The terminology "very powerful" suggests that the authors place high value on the inferential power of the molecular clock analysis, perhaps because it is consistent with the accepted narrative of the outbreak. However, it is interesting that the most likely date *precedes* the arrival of the peacekeepers by more than a week, and the credibility interval extends much deeper into the calendar *before* to the arrival date than it does into the period after arrival. If they had no prior notion of who the guilty party was, the result of this "powerful" technique might direct epidemiological investigators to scrutinize travellers arriving in Haiti during the summer of 2010 to find potential sources of the outbreak!

Conclusions

Both the epidemiological data and molecular clock calculations leave open the possibility that cholera was introduced into Haiti prior to the arrival of the Nepalese contingent, thus weakening the support they provide for the accepted narrative. Clearly the incubation time distribution analysis depends critically on the accuracy of the assertion that the "index patient" experienced onset of the symptoms of cholera on October 12th. A later date for the initial infection event might also be more consistent with the overall narrative regarding the contamination of the Artibonite river by latrine waste because even infection dates as late as October 11th

leave only a few days for the latrines to overflow and for waste to be transported to the alleged dump sites. Nonetheless, the October 12th date is a "discoverable" piece of evidence that appears to have been endorsed by the scientific community.

Similarly, molecular clock estimates depend critically on the assumed rate of mutation. It is quite possible that additional experiments and data refinement could shift the confidence interval to exclude dates earlier than October 8th. But again, in the absence of such refinement the published result becomes potential evidence to throw doubt on the accepted narrative.

Both of these examples point out a more fundamental issue with the current state of microbial forensics as a tool for "resolving controversies" [14]. Many of the scientific studies of the Haiti outbreak play a dual role, satisfying purely academic interests on the one hand, but also published with their forensic case significance in mind. While the degree of explicit advocacy varies from paper to paper, the difference in standards of proof in the two arenas is seldom acknowledged, and perhaps not even understood clearly by the scientists involved. Often ambiguous statements are used to express the significance of findings, their evidential strength, or their degree of certainty, and interpretation is too easily influenced by the prevailing hypothesis and the absence of clearly articulated alternatives. Any findings that do not contradict the favored narrative are taken as support for it. In this situation there is a distinct danger that, under more intense scrutiny such microbial forensic evidence will prove less reliable than some experts anticipate.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Notes and References

[1] Piarroux R, Barrais R, Faucher B, Haus R, Piarroux M, Gaudart J, Magloire R, and Raoult D, "Understanding the Cholera Epidemic in Haiti", Emerging Infectious Diseases 2011; **17**(7) 1161-1167.

[2] Quigley F, "Haiti Strikes Back", Foreign Affairs May 13, 2014.

[3] "Haitians blaming Nepali Nato Units for cholera outbreak", Examiner.com, October 30th, 2010 9:58 pm ET; http://www.examiner.com/foreign-policy-in-national/haitians-blaming-nepali-nato-units-for-cholera-outbreak?cid=parsely#parsely

[4] Georges v. United Nations; Lawsuit filed with the United States District Court, Southern District of New York, October 2013.

[5] Schaefer B, "Haiti Cholera Lawsuit Against the UN: Recommendations for U.S. Policy", The Heritage Foundation Backgrounder No. 2859, November 12, 2013.

[6] Huminer D, Rosenfeld JB, Pitlik SD, "AIDS in the Pre-AIDS Era", Reviews of Infectious Diseases 1987; 9(6): 1102-1108.

[7] Bencic Z and Sinha R, "Cholera Carriers and Circulation of Cholera Vibrios in the Community", International Journal of Epidemiology 1972; 1(1): 13-14.

[8] US Army Corps of Engineers, "Water Resources Assessment of Haiti", August 1999, pp. 2-3; A post-outbreak study of intestinal illness in Haiti found that between 20 and 40% of patients presenting at treatment centers tested positive for cholera: Charles M, et. al., "Importance of Cholera and Other Etiologies of Acute Diarrhea in Post-Earthquake Port-au-Prince, Haiti", Am. J. Trop. Med. Hyg. 2014; 90(3): 511-517.

[9] Cravioto A, Lanata CF, Lantagne DS, and Nair GB, "Final Report of the Independent Panel of Experts on the Cholera Outbreak in Haiti", Available at www.un.org/News/dh/infocus/haiti/UN-cholera-report-final.pdf; Accessed on July 15, 2014.

[10] There is an interesting comment in an article authored by Rumi Chunara , Jason R. Andrews , and John S. Brownstein, "Social and News Media Enable Estimation of Epidemiological Patterns Early in the 2010 Haitian Cholera Outbreak", Am. J. Trop. Med. Hyg. 2012; 86(1): 39–45: "Tweets regarding cholera existed before the start of the cholera outbreak in Haiti"; but they do not specifically say if any of these tweets mention Haiti and cholera simultaneously. Moreover, they do not comment on whether there was tweeting about severe diarrhea in Haiti prior to the outbreak. Considering recent interest in tweet traffic as a surveillance system for influenza, it is possible that an upswing in diarrhea related tweets local to Mirebalais prior to October 8th 2010 would constitute evidence for an earlier introduction.

[11] Piarroux R, "Mission report on the cholera epidemic in Haiti", Unpublished mission report, translated by R.R. Frerichs on January 1, 2011; www.ph.ucla.edu/epi/snow/cholera_haiti_piarrouxreport.html; Accessed on July 15, 2014.

[12] "Update: Cholera Outbreak – Haiti, 2010", Morbidity and Mortality Weekly Report (MWIR) November 19, 2010; 59(45): 1473-1474.

[13] Louise C. Ivers * and David A. Walton, The "First" Case of Cholera in Haiti: Lessons for Global Health; Am. J. Trop. Med. Hyg., 86(1), 2012, pp. 36–38.

[14] Fabini D. Orata, Paul S. Keim, Yan Boucher, "The 2010 Cholera Outbreak in Haiti: How Science Solved a Controversy", PLOS Pathogens 2014; 10(4):e1003967.

[12] Centers for Disease Control, "Acute Watery Diarrhea and Cholera, Pre-Decision Brief for Public Action", February 2010; emergency.cdc.gov/disasters/earth-quakes/Haiti/waterydiarrhea_pre_decision_brief.asp; Accessed on July 14, 2014.

[16] King AA, Ionides EL, Pascual M, and Bouma MJ, "Inapparent infections and cholera dynamics", Nature 2008; 454: 877-881.

[17] Morris GL, "Cholera-Modern Pandemic Disease of Ancient Lineage", Emerging Infectious Diseases 2011; **17**(11): 2099-2104.

[18] World Health Organization (WHO), "Cholera fact sheet", February 2014; www.who.int/mediacentre/factsheets/fs107/en/; Accessed on July 15, 2014.

[19] Azman AS, Rudolph KE, Cummings DAT, Lessler J; "The incubation period of cholera: A systematic review", Journal of Infection 2013; 66: 432 - 438.

[20] Association of Forensic Science Providers, "Standards for the formulation of evaluative forensic science expert opinion", Science and Justice 2009; 49: 161-164.

[21] Katz LS, Petkau A, Beaulaurier J, Tyler S, Antonova ES, Turnsek MA, Guo Y, Wang S, Paxinos EE, Orata F, Gladney LM, Stroika S, Folster JP, Rowe L, Freeman MM, Knox N, Frace M, Boncy J, Graham M, Hammer BK, Boucher Y, Bashir A, Hanage WP, Van Domselaar G, Tarr CL , "Evolutionary Dynamics of Vibrio cholerae O1 following a Single- Source Introduction to Haiti", mBio 2013; 4(4): e00398-13.

Appendix A – A simple stochastic model for cholera patient hospital arrivals

Consider a simple stochastic model to describe the arrival of sick people at a hospital, given that they were infected at a certain time. Assume we infect N people on day 0 (this assumes that there is an person who comes in contact with the local population on or prior to day 0). Of those N people, M will become ill enough to visit the local hospital. Let P_s = probability of severe symptoms. Then M of the N people will be randomly assigned to the set of severe cases with a probability given by the binomial distribution,

$$P(M) = \binom{N}{M} P_s^M (1 - P_s)^{N-M} \quad \text{(A1)}$$

When will each person visit the hospital? This is determined by the incubation time distribution $P_j = P(j|0)$ where j = j^{th} day after infection on day 0. If we represent the number of people arriving at the hospital on day j by n_j, then the distribution of arrival numbers is given by the multinomial distribution:

$$P(n_1, n_2, \dots, n_k) = \frac{M!}{n_1! n_2! \cdots n_k!} P_1^{n_1} P_2^{n_2} \cdots P_k^{n_k} \quad \text{(A2)}$$

Note that $\sum_{j=1}^{k} P_j = 1$ and $\sum_{j=1}^{k} n_j = M$.

The incubation time distribution for cholera has been reviewed in Andrew S. Azman, Kara E. Rudolph, Derek A.T. Cummings, Justin Lessler; "The incubation period of cholera: A systematic review", Journal of Infection 2013; 66: 432 - 438.

Using their data for O1 El Tor Ogawa, we constructed the discrete representation of the incubation time distribution displayed in Fig. A1. The empirical distribution that Azman, et. al. derived has a long tail, so we have taken the discrete distribution out to 21 days and assumed that $P_j = 0$ for j > 21; the residual probability for j >21 has been added to $P_{j=21}$. This distribution is shown in Fig. A1. Fig. A2 is a schematic representation of the procedure represented by equations (A1) and (A2).

In a simulation, we generate realizations of the distribution (A2) and count the number of realizations where:

$n_1 \neq 0$,
$n_1 = 0$, and $n_2 \neq 0$,
$n_1 = n_2 = 0$, and $n_3 \neq 0$,
$n_1 = n_2 = n_3 = 0$, and $n_4 \neq 0$, etc.

and make a histogram representing the probability that no patient shows up until the j^{th} day after infection. In practice we run this simulation 10,000 times to obtain a sample of the distribution for a chosen N and P_s. We then average 10 simulations with independent random seeds to obtain the final distribution P(K|0,N)

representing the probability that at least one patient shows up at a hospital on day K if he is one of N victims that are infected on day 0.

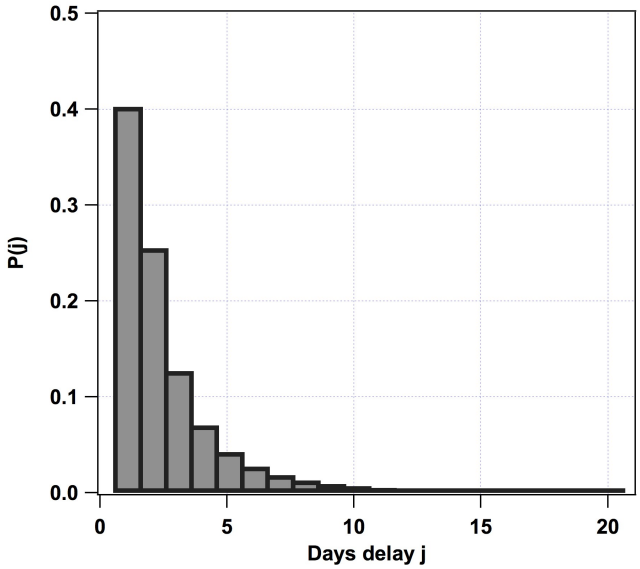

Figure A1. Discretized representation of the incubation time distribution for cholera; derived from Azman, et. al. It is assumed that $P_j = 0$ for $j > 21$ days, and day 21 contains the residual probability associated with $j > 21$.

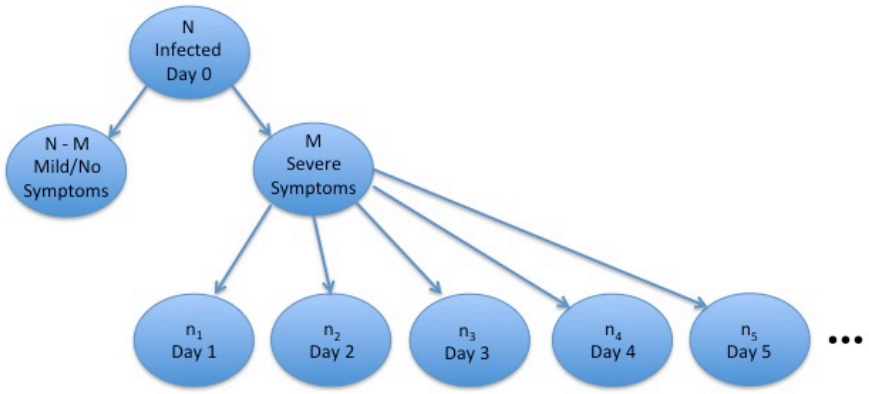

Figure A2. Stochastic model for determining probability of … Each arrow represents a random assignment according to equations (A1) and (A2) .

Appendix B. Calculating the posterior probability

Consider the joint probability distribution for infection of N persons on day J, and first patients showing up on day J+K:

$$P(J, J+K, N) = P(J, N|J+K)P(J+K) = P(J+K|J, N)P(J, N) \quad \text{(B1)}$$

In our discrete event model we imagine this process to take place on a very long chain of bins, each bin corresponding to one day. We pick a bin (day) at random, and place N infected persons in it. We then partition the N persons according to equation (A1) to produce M "critical" patients who are ill enough to seek medical care. These are then partitioned among the bins to the right of bin J according to equation (A2). The first bin that contains at least one patient is the bin labeled J+K, K = 1, 2, 3, … . Note that we treat all bins as equivalent with equal probability of placing the original N infected persons in any bin J.

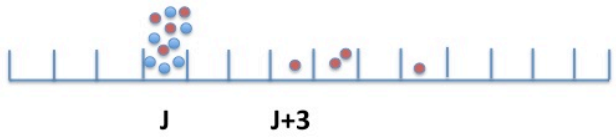

Figure B1. Representation of the stochastic model for cholera. Eleven infected persons are created in bin (day) J, and four (red) become sufficiently ill to require medical attention. The first patient shows up at a hospital on day J + 3.

Let $J + K = Q$. We can then write the joint probability as: $P(Q - K, Q, N)$. Note that in this expression we should obtain the same quantity whether we arbitrarily call the first bin to contain sick patients Q = 0 or call the bin containing the set of N infectees J = 0. Thus:

$$P(\mathbf{0}, \boldsymbol{K}, N) = P(-\boldsymbol{K}, \mathbf{0}, N) \quad \text{(B2)}$$

where it is understood that the ordering of the variables always follows the convention that the first variable (red) refers to the bin of N infectees and the second (blue) to the bin containing the first patient(s). We can then write each joint probability in terms of conditional and prior probabilities:

$$P(\boldsymbol{K}|\mathbf{0}, N)P(\mathbf{0}, N) = P(-\boldsymbol{K}, N|\mathbf{0})P(\mathbf{0}) \quad \text{(B3)}$$

The variable N is conditionally independent of the particular bin in which the victims are placed, so P(**0**,N) = P(**0**)P(N). Because all bins are equivalent, and the convention chosen for assigning a bin the index 0 should not affect probability assignments, the prior probabilities on both sides of the equation are equal: $P(\mathbf{0}) = P(\mathbf{0})$. Thus, we can write:

$$P(\boldsymbol{K}|\mathbf{0}, N)P(N) = P(-\boldsymbol{K}, N|\mathbf{0}) \quad \text{(B4)}$$

The probability on the left-hand side is the probability of observing our first patient(s) in bin K, given that N infected persons were created in bin 0. The right-hand side is the probability that our original batch of N patients was created in bin $-K$, given that our first patient(s) was/were observed in bin 0. We can use this equation in two ways. First, we can regard N as given, so that P(N) = 1 for that value and zero for all other values. In this "fixed N frame" P(**-K**,N|**0**) depends parametrically on N. We can also regard N as a random variable and average P(**-K**,N|**0**) over N:

$$P(-\boldsymbol{K}, \boldsymbol{0}) = \sum_{N=0}^{\infty} P(\boldsymbol{K}|\boldsymbol{0}, N)P(N) \qquad \text{(B5)}$$

A simple model might posit an average rate of infection μ (per bin) with a Poisson distribution of possible infection events:

$$P(N) = \frac{\mu^N e^{-\mu}}{N!} \qquad \text{(B6)}$$

Equations (B4) or (B5) give the probability that N persons were infected on day –K given that we observe the first patient(s) on day 0 in terms of the probability of observing our first patient(s) on day K, given that N persons were infected on day 0. We can use the procedure outlined in Appendix A to calculate P(K|0,N).